

AD A 073796

12

LEVEL

MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION

BY

T. W. ANDERSON and I. OLKIN

TECHNICAL REPORT NO. 38

JULY 1979

PREPARED UNDER CONTRACT

N00014-75-C-0442 (NR-042-034)

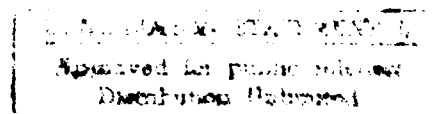
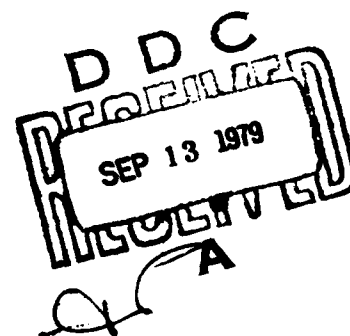
FOR THE OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA



79 09 13 035

DDC FILE COPY

MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION

by

11 T. W./ANDERSON and I. OLKIN

1 TECHNICAL REPORT NO. 38

11 JUL 1979

13 PREPARED UNDER CONTRACT N00014-75-C-0442, NSR-42-034

(NR-042-034)

14 OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.

Also issued as Technical Report No. 144 under National Science Foundation
Grant MCS 78-07736, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

332 580

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION

by

T. W. Anderson and I. Olkin
Stanford University

ABSTRACT

This paper provides an exposition of several alternative techniques used to obtain maximum likelihood estimators for the parameters of a multivariate normal distribution. In particular, matrix differentiation, matrix transformations and induction are treated. These techniques are used to derive the maximum likelihood estimators of the covariances of a Wishart distribution, of the covariances when there are missing observations, and of the means under a rank constraint. Although the paper is mainly expository, some of the proofs are new.

Key words: multivariate normal distribution; Wishart distribution; maximum likelihood estimation; maximization subject to a rank constraint; matrix transformations; matrix differentiation; multivariate inequalities.

Maximum Likelihood Estimation of the Parameters
of a Multivariate Normal Distribution

1. Introduction.

Consider the maximum likelihood estimation of the mean μ and covariance matrix Σ of a normal p -variate distribution based on observations x_1, \dots, x_N . The maximum likelihood estimator of the mean is the sample mean \bar{x} , and the logarithm of the concentrated likelihood is $-\frac{1}{2} pN \log 2\pi + \frac{1}{2} N$ times

$$(1) \quad f(G) = -\log|G| - \text{tr } G^{-1}V$$

or

$$(2) \quad g(H) = \log|H| - \text{tr } H V ,$$

where

$$(3) \quad V = \frac{1}{N} \sum_{\alpha=1}^N (x_{\alpha} - \bar{x})(x_{\alpha} - \bar{x})'$$

is positive definite, G represents the covariance matrix Σ and H the inverse Σ^{-1} . The problem is to maximize (1) with respect to positive definite G . Because H is a one-to-one transformation of G , it is equivalent to maximize (2) with respect to positive definite H . (This is Lemma 3.2.2 of Anderson (1958).)

We use the notation $A > 0$ to mean that the symmetric matrix A is positive definite.

We shall describe three alternative approaches: differentiation, matrix transformations, and induction, each of which can be applied to $f(G)$ or $g(H)$. It is expected that these techniques will be useful in other problems of maximization. Three examples are given to illustrate the ideas.

Remark. It is somewhat surprising that an early reference to the fact that the sample covariance matrix is the maximum likelihood estimator of Σ is elusive. The general result is implicit in the work of Wilks (1932, p. 476) dealing with likelihood ratio tests.

2. The Method of Differentiation.

The functions $f(G)$ and $g(H)$ go to $-\infty$ as G or H approaches the boundary of positive definite matrices or as one or more elements increases without bound. From the facts that $\log|H|$ is concave (see Bellman (1970), p. 128) and $\text{tr } H$ is linear, it follows that $g(H)$ is concave in $H > 0$, so that a maximum exists, and it is unique. The function $f(G)$ is neither concave nor convex. However, since there is only one solution to the derivative equations, these equations yield the maximum.

To obtain the derivative equations we use the differential form

$$(3) \quad d[f(G)]_{ij} = (2 - \delta_{ij}) \left\{ -\frac{G_{ij}}{|G|} dg_{ij} + (\text{tr } G^{-1} E_{ij} G^{-1} V) dg_{ij} \right\} = 0, \quad i \leq j,$$

where δ_{ij} is the Kronecker delta, E_{ij} is a matrix with 1 in the (i,j) th position and 0 elsewhere, and G_{ij} is the cofactor of g_{ij} . This yields the matrix equation

$$(4) \quad -G^{-1} + G^{-1} V G^{-1} = 0,$$

with the unique solution $\hat{G} = V$.

Similarly, using $g(H)$,

$$(5) \quad d[g(H)]_{ij} = (2 - \delta_{ij}) \left\{ \frac{H_{ij}}{|H|} dh_{ij} - (\text{tr } E_{ij} V) dh_{ij} \right\} = 0, \quad i \leq j,$$

which yields the equation

$$(6) \quad H^{-1} - V = 0,$$

with the unique solution $\hat{H} = V^{-1}$.

In essence, this approach applied to $f(G)$ has been discussed by Smith (1978); the approach applied to $g(H)$ was used by Anderson (1958). For a more detailed discussion of differentials see Deemer and Olkin (1951), or Anderson (1958) p. 310.

3. The Method of Matrix Transformations.

The functions $f(G)$ and $g(H)$ can be written in canonical forms. For any matrix C such that $CC' = V$, let

$$\tilde{G} = C^{-1}GC'^{-1}, \quad \tilde{H} = C'HC.$$

Then (1) and (2) yield

$$f(G) - \log |V| = -\log |\tilde{G}| - \text{tr } \tilde{G}^{-1},$$

$$g(H) - \log |V| = \log |\tilde{H}| - \text{tr } \tilde{H}.$$

Hence we shall take $V = I$ and drop the tilde.

We use three well-known representations for a positive definite matrix, namely,

$$(i) \quad H = TT',$$

where T is upper (or lower) triangular;

$$(ii) \quad H = D R D,$$

where $D = \text{diag}(\sqrt{h_{11}}, \dots, \sqrt{h_{pp}})$, and $R = (r_{ij})$ is a correlation matrix, (i.e., $r_{ii} = 1$);

$$(iii) \quad H = \Gamma D_d \Gamma',$$

where Γ is orthogonal, $D_d = \text{diag}(d_1, \dots, d_p)$ and d_1, \dots, d_p are the characteristic roots of H .

In each case the problem reduces to

$$(6) \quad \underset{z > 0}{\text{Max}} \quad (\log z - z).$$

The function $\log z - z$ is concave and has the unique maximum of -1 at $z = 1$.

3.1. Transformation to Rectangular Coordinates: $H = TT'$. Now the maximization of $g(H)$ becomes

$$(7) \quad \underset{\substack{t_{11}, t_{ij} \\ i < j}}{\text{Max}} \quad \left\{ \sum_i (\log t_{ii}^2 - t_{ii}^2) - \sum_{i < j} t_{ij}^2 \right\}.$$

Clearly, the maximum over t_{ij} , $i < j$ occurs at $t_{ij} = 0$, so that

(7) reduces to a sum of terms like (6).

3.2. Transformation to a Correlation Matrix: $H = D R D$. Now the maximization of $g(H)$ becomes

$$(8) \quad \underset{\substack{h_{ii} > 0 \\ R > 0}}{\text{Max}} \left\{ \sum_1 (\log h_{ii} - h_{ii}) + \log |R| \right\}.$$

Since R is a correlation matrix, by Hadamard's determinant inequality $|R| \leq \prod r_{ii} = 1$, with equality for $R = I$. Alternatively, we can write $R = UU'$, where U is upper triangular with $\sum_{j=1}^1 u_{ij}^2 = 1$; consequently, $|R| = \prod u_{ii}^2 \leq 1$. Now (8) reduces to a sum of terms like (6).

3.3. Transformation to Characteristic Roots: $H = \Gamma D_d \Gamma'$. Now the maximization of $g(H)$ becomes

$$\underset{d_i > 0}{\text{Max}} \left\{ \sum_1 (\log d_i - d_i) \right\},$$

which is of the form (6). The transformation 3.3 has been suggested by Anderson (1958) Problem 4, Chapter 3, and has been used by Watson (1964); it is the essence of the method of Khatri (1979) and Tamhane (1979).

4. The Method of Induction.

Write

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & h_{22} \end{pmatrix}, \quad H_{11}: p-1 \times p-1;$$

then the maximization of $g(H)$ becomes

$$\begin{array}{c} \text{Max} \\ H_{11}, H_{12}, h_{22} \\ H > 0 \end{array} \{ (\log |H_{11}| - \text{tr } H_{11}) + \log(h_{22} - H_{21} H_{11}^{-1} H_{12}) - h_{22} \}.$$

The maximum with respect to H_{21} is achieved at $H_{21} = 0$ for any H_{11} and h_{22} , which yields

$$\begin{array}{c} \text{Max} \\ H_{11} > 0, \\ h_{22} > 0 \end{array} \{ (\log |H_{11}| - \text{tr } H_{11}) + (\log h_{22} - h_{22}) \}.$$

The result now follows from the inductive hypothesis.

5. Other Methods.

If $f(G)$ and $g(H)$ are not put into canonical form several other methods may be employed. There is a nonsingular matrix Q such that

$$V = QQ', \quad G = QD_{\lambda}Q',$$

where D_{λ} is a diagonal matrix with diagonal elements the roots of $|G - \lambda V| = 0$. Then

$$f(G) = -2 \log |Q| - \sum_{i=1}^p (\log \lambda_i + \frac{1}{\lambda_i}),$$

which is maximized for $\hat{\lambda}_i = 1$; that is, $\hat{D}_{\lambda} = I$, $\hat{G} = V$.

Another method invokes a theorem of von Neumann (1937):

$$\text{tr } H V \geq \sum_{i=1}^p \eta_i v_{p-i+1}$$

where $\eta_1 > \dots > \eta_p$ are the ordered characteristic roots of H and $v_1 > \dots > v_p$ are the ordered characteristic roots of B ; equality is attained when the characteristic vectors of H are identical to those of V . We have

$$g(H) \leq \sum_{i=1}^p (\log \eta_i - \eta_i v_{p-i+1})$$

with equality when the characteristic vectors of H are identical to those of V . Then $(\log \eta_i - \eta_i v_{p-i+1})$ is maximized at $\eta_i = 1/v_{p-i+1}$, $i = 1, \dots, p$. Thus $g(H)$ is maximized at $\hat{H} = V^{-1}$. This method is used by Theobald (1975).

6. A Modified Model with Missing or Additional Observations.

Suppose a sample of size N is observed from a p -variate normal distribution with covariance matrix Σ , and an additional sample of size M is observed on the first k (out of p) variates. Alternatively, this model can be viewed as a sample of size $N+M$ from a p -variate normal distribution, where M observations on the last $p-k$ (out of p) variates are missing. This problem was considered by Anderson (1957) and in another context by Olkin and Sylvan (1977).

The problem now is to

$$(9) \quad \max_{G>0} \{-N \log |G| - \text{tr } G^{-1}V - M \log |G_{11}| - \text{tr } G_{11}^{-1}W\},$$

where $G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$, $G_{11}: k \times k$, $V: p \times p$ and $W: k \times k$ are positive definite.

6.1. Reduction to a Canonical Form. In terms of $H = G^{-1}$, the problem becomes

$$\begin{aligned}
 (10) \quad & \text{Max}_{H>0} \{N \log |H| - \text{tr } H V + M \log |H_{11} - H_{12} H_{22}^{-1} H_{21}| - \text{tr}(H_{11} - H_{12} H_{22}^{-1} H_{21}) W\} \\
 & = \text{Max}_{H>0} \{N \log |H_{22}| + (N + M) \log |H_{11} - H_{12} H_{22}^{-1} H_{21}| - \text{tr } P_{11} V_{11} - \text{tr } H_{12} V_{21} \\
 & \quad - \text{tr } H_{21} V_{12} - \text{tr } H_{22} V_{22} - \text{tr}(H_{11} - H_{12} H_{22}^{-1} H_{21}) W\} .
 \end{aligned}$$

6.2. Transformation.

One can think of this likelihood as composed of the marginal likelihood of the $N + M$ observations on the k variables and the conditional likelihood function of the N observations on the $p - k$ variables conditional on the N observations on the k variables. The appropriate parameters are

$$H_{11} - H_{12} H_{22}^{-1} H_{21} = J, \quad H_{22}^{-1} H_{21} = K, \quad H_{22} = H_{22} .$$

(If $H = G^{-1}$, then $J = G_{11}^{-1}$, $K = -G_{21} G_{11}^{-1}$, $H_{22} = (G_{22} - G_{21} G_{11}^{-1} G_{12})^{-1}$.)

The corresponding transformation of the covariance matrix of N observations is

$$V_{22} - V_{21} V_{11}^{-1} V_{12} = F, \quad V_{21} V_{11}^{-1} = E, \quad V_{11} = V_{11} .$$

Then (10) is transformed to

$$(11) \quad \text{Max}_{H_{22} > 0, J > 0, K} \{ N \log H_{22} - \text{tr } H_{22} F + (N+M) \log |J| - \text{tr } J(V_{11} + W) \\ - \text{tr } H_{22}(K+E)V_{11}(K+E)' \}.$$

The last term is nonpositive because H_{22} and V_{11} are positive definite; its maximum of 0 occurs at $\hat{K} = -E = -V_{21}V_{11}^{-1}$. Then (11) becomes the sum of two terms like (2). The maximum over $H_{22} > 0$ and $J > 0$ occurs at $\hat{H}_{22} = NF^{-1}$ and $\hat{J} = (N+M)(V_{11}+W)^{-1}$.

It might be noted that this last analysis is applicable to the following problem. Given a matrix X with regression KZ and covariance matrix H_{22}^{-1} , the log likelihood function is (11) with $J = 0$, $E = X'Z(Z'Z)^{-1}$ and $F = X'X - EZ'ZE'$.

6.3. Differentiation.

We now solve (9) by differentiation; this alternative may have some intrinsic interest. Using differential forms, we obtain

$$(2-\delta_{ij}) \left\{ \frac{G_{ij}^*}{|G|} dg_{ij} + (\text{tr } G^{-1} E_{ij} G^{-1} V) dg_{ij} - M \begin{pmatrix} \frac{A_{ij}^*}{|G_{11}|} da_{ij} & 0 \\ 0 & 0 \end{pmatrix} + da_{ij} \text{tr} \begin{pmatrix} G_{11}^{-1} E_{ij} G_{11}^{-1} W & 0 \\ 0 & 0 \end{pmatrix} \right\} = 0,$$

$i \leq j$, where $A \equiv G_{11}$ and A_{ij}^* is the cofactor of a_{ij} in A . This yields the matrix equation

$$(12) \quad -NG^{-1} + G^{-1}VG^{-1} - M \begin{pmatrix} G_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} G_{11}^{-1} W G_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

Pre- and post-multiplication by G in (12) yields

$$(13) \quad -NG + V - M G \begin{pmatrix} G_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} G + G \begin{pmatrix} G_{11}^{-1} W G_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} G = 0 ,$$

which simplifies to the set of equations

$$(14a) \quad -NG_{11} + V_{11} - MG_{11} + W = 0 ,$$

$$(14b) \quad -NG_{12} + V_{12} - MG_{12} + WG_{11}^{-1}G_{12} = 0 ,$$

$$(14c) \quad -NG_{22} + V_{22} - MG_{21}G_{11}^{-1}G_{12} + G_{21}G_{11}^{-1}WG_{11}^{-1}G_{12} = 0 .$$

The equations (14) can be solved in sequence to yield

$$(15a) \quad \hat{G}_{11} = \frac{1}{N+M} (V_{11} + W) ,$$

$$(15b) \quad \hat{G}_{12} = \frac{1}{N+M} (V+W)V_{11}^{-1}V_{12} ,$$

$$(15c) \quad \hat{G}_{22} = \frac{1}{N} V_{22} + \frac{1}{(N+M)} V_{21}V_{11}^{-1}WV_{11}^{-1}V_{12} - \frac{M}{N(N+M)} V_{21}V_{11}^{-1}V_{12} .$$

7. A Problem of Rank.

Let the N columns of X and the M columns of Y be independently distributed according to a p -variate normal distribution with covariance matrix Σ and $EX = 0$, $EY = \phi$, where ϕ is of rank $r \leq p \leq M$. This model is considered by Anderson (1951).

To obtain the maximum likelihood estimators of Σ and Φ we start with the likelihood function

$$(16) \quad c |\Sigma|^{-\frac{1}{2}(N+M)} \exp\left\{-\frac{1}{2} \text{tr } \Sigma^{-1}[X X' + (Y-\Phi)(Y-\Phi)']\right\},$$

where c is a normalizing constant. From Sections 1, 2, and 3, the maximum of (16) with respect to Σ (for fixed Φ) is

$$\hat{\Sigma} = [X X' + (Y - \Phi)(Y - \Phi)'] / (N+M),$$

so that we need to determine

$$(17) \quad \begin{aligned} \min_{\Phi} |\hat{\Sigma}| &= \min_{\Phi} |X X' + (Y - \Phi)(Y - \Phi)'| \\ &= \min_{\Phi} |X X'| |I + (X X')^{-\frac{1}{2}}(Y - \Phi)(Y - \Phi)'(X X')^{-\frac{1}{2}}|. \end{aligned}$$

To simplify notation, write

$$(18) \quad \tilde{Y} = (X X')^{-\frac{1}{2}} Y, \quad \tilde{\Phi} = (X X')^{-\frac{1}{2}} \Phi,$$

then (17) becomes (except for the term $|X X'|$)

$$(19) \quad \min_{\Phi} |I + (\tilde{Y} - \tilde{\Phi})(\tilde{Y} - \tilde{\Phi})'|.$$

Since $\tilde{\Phi}: p \times M$ is of rank r , we can write

$$(20) \quad \tilde{\Phi} = (T_1 \ T_2) \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix} = T_1 \Lambda_1,$$

where $(T_1 \ T_2): p \times p$, $T_1: p \times r$; $\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}: M \times M$, $\Delta_1: r \times M$ and Δ is orthogonal. In the representation (20), $\tilde{\Phi} \tilde{\Phi}' = T_1 T_1'$. If $T_1 = \begin{pmatrix} T_{11} \\ T_{21} \end{pmatrix}$, where $T_{11}: r \times r$ is lower triangular with positive diagonal elements, then the representation $\tilde{\Phi} = T_1 \Delta_1$ is unique.

In (20) note that

$$\begin{aligned}
 (21) \quad & |I_p + (\tilde{Y} - \Phi)(\tilde{Y} - \Phi)'| \\
 &= |I_p + \tilde{Y} \tilde{Y}' + T_1 T_1' - T_1 \Delta_1 \tilde{Y}' - \tilde{Y} \Delta_1' T_1'| \\
 &= |(I_p + \tilde{Y} \tilde{Y}' - \tilde{Y} \Delta_1' \Delta_1 \tilde{Y}') + (T_1 - \tilde{Y} \Delta_1')(T_1 - \tilde{Y} \Delta_1')'| \\
 &= |I_p + \tilde{Y} \Delta_2' \Delta_2 \tilde{Y}' + (T_1 - \tilde{Y} \Delta_1')(T_1 - \tilde{Y} \Delta_1')'|.
 \end{aligned}$$

Since $I_p + \tilde{Y} \Delta_2' \Delta_2 \tilde{Y}'$ is positive definite, the minimum of (21) over T_1 is achieved at $\hat{T}_1 = \tilde{Y} \Delta_1$, in which case we need to determine

$$\begin{aligned}
 (22) \quad & \min_{\Delta_2} |I_p + \tilde{Y} \Delta_2' \Delta_2 \tilde{Y}'| = \min_{\Delta_2} |I_{M-r} + \Delta_2 \tilde{Y}' \tilde{Y} \Delta_2'| \\
 &= \min_{\Delta_2} |\Delta_2 (I_{M-r} + \tilde{Y}' \tilde{Y}) \Delta_2'|.
 \end{aligned}$$

The minimum of (22) is $\prod_{r+1}^p \lambda_i$, where $\lambda_1 > \dots > \lambda_p$ are the ordered characteristic roots of $I + \tilde{Y}' \tilde{Y} = I + Y'(XX')^{-1}Y$, that is, the roots of $|(X'X + Y'Y) - \lambda X'X| = 0$. (See, e.g., Bellman (1970),

Theorem 10, p. 132.) The minimum is attained at

$$\Delta_2 = (c_{r+1}, \dots, c_p),$$

where c_i is the vector satisfying $(I + Y'(XX')^{-1}Y)c_i = \lambda_i c_i$ and $c_i' c_i = 1$.

Combining our results we obtain as the maximized likelihood

$$(23) \quad \frac{c e^{-\frac{1}{2}p} (N+M)^{\frac{1}{2}p(N+M)}}{|XX'|^{-\frac{1}{2}(N+M)} \left(\prod_{i=r+1}^p \lambda_i \right)^{\frac{1}{2}(N+M)}}.$$

The model (16) is considered by Healy (1979). His procedure is to first make a series of transformations motivated by the rank condition on Φ , which transforms the model to a canonical form, and then to carry out the maximizations. However, by first maximizing with respect to the covariances (as in the above derivation), the required transformations (and proof) become considerably simpler.

REFERENCES

- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist., Vol. 22, pp. 327-351.
- Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. J. Amer. Statist. Assoc., 52, 200-203.
- Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc.: New York.
- Bellman, R. (1970). Introduction to Matrix Analysis. 2nd edition. McGraw Hill Book Co.: New York.
- Deemer, W., Jr. and Olkin, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. Biometrika, Vol. 38, 345-367.
- Healy, J. D. (1979). Maximum likelihood estimation of a linear functional relationship. J. Multivariate Analysis, Vol. 9, in press.
- Khatri, C. G. (1979). Letter to the Editor. The American Statistician, 33, 92.
- von Neumann, J. (1937). Some matrix inequalities and metrization of matric-space. Tomsk University Review, 1, 286-300. Reprinted in John Von Neumann Collected Works (1962), Vol. 4 (ed. by A. H. Taub). Pergamon Press: New York.
- Olkin, I. and Sylvan, M. (1977). Correlational analysis when some variances and covariances are known. Multivariate Analysis, IV (ed. by P. R. Krishnaiah), pp. 175-191. North-Holland Publishing Co.
- Smith, D. W. (1978). A simplified approach to the maximum likelihood estimation of the covariance matrix. The American Statistician, 32, 28-29.
- Tamhane, A. C. (1979). Letter to the Editor. The American Statistician, 33, 92-93.
- Theobald, C. M. (1975). An inequality with application to multivariate analysis. Biometrika, 62, 461-466.
- Watson, G. S. (1964). A note on maximum likelihood. Sankhya A, 26, 303-304.
- Wilks, S. S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.
2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.
3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.
4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.
5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.
6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.
7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.
8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.
9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.
10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.
11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.
12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.
13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.
14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.
15. "Maximum Likelihood Estimation of Parameter of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.
16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations,"
Steffen L. Lauritzen, May, 1974.
18. "On the Interrelationships among Sufficiency, Total Sufficiency and
Some Related Concepts," Steffen L. Lauritzen, June, 1974.
19. "Statistical Inference for Multiply Truncated Power Series Distributions,"
T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models
in the Time and Frequency Domains," T. W. Anderson, June 1975.
21. "Asymptotic Properties of Some Estimators in Moving Average Models,"
Raul Pedro Mentz, September 8, 1975.
22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
Mituaki Huzii, February 1976.
23. "Estimating Means when Some Observations are Classified by Linear
Discriminant Function," Chien-Pai Han, April 1976.
24. "Panels and Time Series Analysis: Markov Chains and Autoregressive
Processes," T. W. Anderson, July 1976.
25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson,
September 1976.
26. "The Recurrence Classification of Risk and Storage Processes,"
J. Michael Harrison and Sidney I. Resnick, September 1976.
27. "The Generalized Variance of a Stationary Autoregressive Process,"
T. W. Anderson and Raul P. Mentz, October 1976.
28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
Javad Behboodian, October 1976.
29. "Identification of Parameters by the Distribution of a Maximum
Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.
30. "Discrimination Between Stationary Gaussian Processes, Large Sample
Results," Will Gersch, January 1977.
31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
Christine M. Waternaux, October 1977.
32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a
Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

TECHNICAL REPORTS (continued)

33. "Canonical Correlations with Respect to a Complex Structure," Steen A. Andersson, July 1978.
34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and I. Olkin, July 1978.
35. "Maximum likelihood Estimation for Vector Autoregressive Moving Average Models," T. W. Anderson, July 1978.
36. "Maximum likelihood Estimation of the Covariances of the Vector Moving Average Models in the Time and Frequency Domains," F. Ahrabi, August 1978.
37. "Efficient Estimation of a Model with an Autoregressive Signal with White Noise," Y. Hosoya, March 1979.
38. "Maximum Likelihood Estimation of the Parameters of a Multivariate Normal Distribution," T.W. Anderson and I. Olkin, July 1979.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 38	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A MULTIVARIATE NORMAL DISTRIBUTION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. ANDERSON and I. OLKIN		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE JULY 1979
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Issued also as Technical Report No. 144 under National Science Foundation Grant MCS 78-07736, Department of Statistics, Stanford University.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multivariate normal distribution; Wishart distribution; maximum likelihood estimation; maximization subject to a rank constraint; matrix transforma- tions; matrix differentiation; multivariate inequalities.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) SEE REVERSE SIDE.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT.

This paper provides an exposition of several alternative techniques used to obtain maximum likelihood estimators for the parameters of a multivariate normal distribution. In particular, matrix differentiation, matrix transformations and induction are treated. These techniques are used to derive the maximum likelihood estimators of the covariances of a Wishart distribution, of the covariances when there are missing observations, and of the means under a rank constraint. Although the paper is mainly expository, some of the proofs are new.

38/144

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)